

Modelling the Variability of the Wind Energy Resource on Monthly and Seasonal Timescales

Bastien Alonzo^{1,2}, Hans-Kristian Ringkjøb^{1,2,4}, Benedicte Jourdier^{1,3},
Philippe Drobinski¹, Riwal Plougonven¹, Peter Tankov²

Abstract

An avenue for modelling part of the long-term variability of the wind energy resource from knowledge of the large-scale state of the atmosphere is investigated. The timescales considered are monthly to seasonal, and the focus is on France and its vicinity. On such timescales, one may obtain information on likely surface winds from the large-scale state of the atmosphere, determining for instance the most likely paths for storms impinging on Europe. In a first part, we reconstruct surface wind distributions on monthly and seasonal timescales from the knowledge of the large-scale state of the atmosphere, which is summarized using a principal components analysis. We then apply a multi-polynomial regression to model surface wind speed distributions in the parametric context of the Weibull distribution. Several methods are tested for the reconstruction of the parameters of the Weibull distribution, and some of them show good performance. This proves that there is a significant potential for information in the relation between the synoptic circulation and the surface wind speed. In the second part of the paper, the knowledge obtained on the relationship between the large-scale situation of the atmosphere and surface wind speeds is used in an attempt to forecast wind speeds distributions on a monthly horizon. The forecast results are promising but they also indicate that the Numerical Weather Prediction seasonal forecasts on which they are based, are not yet mature enough to

Email address: `bastien.alonzo@lmd.polytechnique.fr` (Bastien Alonzo)

¹IPSL/LMD, CNRS, Ecole Polytechnique, Université de Paris-Saclay, Palaiseau, France

²Laboratoire de Probabilités et Modèles Aléatoires, Université Paris Diderot - Paris 7, Paris, France.

³EDF, R&D, Chatou, France

⁴Geophysical Institute, University of Bergen, Bergen, Norway

provide reliable information for timescales exceeding one month.

Keywords: Seasonal modelling, Wind distribution, Variability, large-scale circulation, Forecasts, Wind energy

1. Introduction

Owing to a well-established technology and the ever stronger push towards replacing fossil fuels with clean renewable power, wind energy has seen a dramatic growth in the recent years. According to the European Wind Energy Association, about 12.8 GW of wind power was installed in the European Union (EU) in 2015, bringing EUs total installed capacity to 141.6 GW. This corresponds to an electricity generation sufficient to cover 11.4 % of the EUs electricity consumption during an average year [1].

With the growing importance of wind energy, the interest and demand for forecasts of the wind speed near the surface has seen a major boost. Numerous methods exist for forecasting the wind speeds at different forecast horizons implying different applications [2, 3]. Many studies focus on the short-term scale ranging from several minutes to 1 day [4, 5, 6]. Medium-term forecast methods, ranging from several days up to 10 days, have also been well investigated [7, 8, 9]. On much longer timescales and with very different implications and motivations, the impact of climate change on wind speeds has also been addressed [10, 11, 12].

By contrast, the intermediate timescale ranging from one month to a season (hereafter referred to as long-term) has received only little attention. Monthly and seasonal forecasts can be very useful for example in maintenance planning, financial estimates and predictions of electricity generation for network management. Some studies showed good results in forecasting the monthly mean wind speed at several observation sites by using Artificial Neural Network models (ANN) [13, 14], giving an accurate trend of the wind speed at the yearly horizon, but a limited information on the wind variability at higher frequency. Other authors forecasted daily mean wind speed at the seasonal scale using ANN [15, 16, 17] allowing to gather more informations on the wind variability inside a given season and which would allow to evaluate the energy production. The ANN output is a predicted wind time serie. They calculate the error regarding the real wind speed and compare the results to other ANN [15, 16] or other statistical methods namely ARIMA models [17]. As ANN behaves like black box which we feed with data, the results are

33 difficult to explain physically. Moreover, each methods focuses on different
 34 observation sites giving a limited idea of the spatial variations of the method
 35 performance. Even though there are very few works on seasonal forecasts
 36 of wind speeds, seasonal forecasting of other meteorological quantities is a
 37 popular research topic with continuous improvement. For example, there
 38 have been many works on seasonal forecasts of recurrent oscillating patterns
 39 in the atmosphere, such as the El Nino [18, 19].

40 This paper focuses on modelling the wind variability on the long-term
 41 timescale and makes an attempt of long-term wind speed distribution fore-
 42 casting. The method proposed in this work aims to use the information found
 43 in the large-scale configuration of the atmosphere in order to reconstruct ex-
 44 pected distribution of surface winds. This paper answers some questions that
 45 arise from this topic :

- 46 • How much information on the monthly or seasonal distribution of sur-
 47 face winds can we obtain from knowledge of only the large-scale state
 48 of the atmosphere?
- 49 • Are the proposed methods performing better than the climatology in
 50 reproducing the surface wind speed distribution, and in estimating the
 51 electricity generation?
- 52 • Do seasonal forecasts from an operational center of weather production
 53 contain relevant information for an attempt of forecasting wind speed
 54 distributions and electricity generation?

55 To address these questions in a consistent framework, we use data from
 56 the European Center for Medium-Range Weather Forecasts (ECMWF). In-
 57 deed, surface winds from ECMWF reanalysis have been shown to well re-
 58 produce the observed surface winds in France [20]. Using reanalysis data
 59 allows a better investigation of the statistical relation between local surface
 60 winds and the large-scale circulation variability, especially because they pro-
 61 vide a continuous description of surface wind speed over a wide domain and
 62 over long time period. We focus on France and its vicinity not only because
 63 the reanalyzed winds had been assessed there, but also because France has
 64 a significant wind energy potential and interestingly includes regions with
 65 different wind regimes. In Northern France the wind energy potential stems
 66 from the storm tracks, whereas local orographic effects and channeling play
 67 a major role in strong wind events of Southern France [21].

68 In the first part of this paper, the data and methodology used to link
69 the large scale circulation with the surface wind speed and to reconstruct
70 its monthly/seasonal distributions is described. Then, the performance of
71 the proposed methods is evaluated by comparing their results to the clima-
72 tology distributions. The performance is evaluated in terms of recontructed
73 electricity generation as well. In the last part of the paper, an attempt in
74 forecasting wind speed distributions and electricity generation is discussed.

75 2. Data and Methods

76 2.1. Data

77 *ERA-Interim reanalysis.* Wind speed, geopotential height at 500hPa (Z500) and
78 Mean Sea Level Pressure (MSLP) are collected from ERA-Interim reanal-
79 ysis (ERA-Interim, [22]) with a time-step of six hours during 35 years between
80 01/01/1979 and 12/31/2013, and then averaged to daily data. The horizon-
81 tal resolution of ERA-Interim is 0.75° in latitude and longitude. Z500 and MSLP
82 span the North Atlantic and European grid (20°N to 80°N and 90°W to
83 40°E), and the surface wind speeds are obtained for a domain encompassing
84 France (40.5°N to 52.5°N and -6.75°W to 10.5°E).

85 The ERA-Interim reanalysis data act as the reference data for wind speed. B.
86 Jourdain [20] showed that the ERA-Interim reanalysis has a good skill for
87 wind speeds in France, and is the best in comparison to two other reanalyses:
88 MERRA and the NCEP/NCAR. To reconstruct the distribution of the wind,
89 a 20 years calibration period, on which we train our methods, has been
90 defined from 1 January 1979 to 31 December 1998. Then a validation period
91 lasting 15 years from 01 January 1999 to 31 December 2013 follows.

92 *ECMWF Forecasts.* In the forecast section, the full 35 years period of ERA-
93 Interim is used as a calibration period, while the period of forecast is always of 3
94 months, permitting to predict either monthly or seasonal distribution of the
95 surface wind speed. We retrieve twelve seasonal forecast sets of ECMWFs
96 numerical weather prediction model [23], from the years 2012, 2013 and 2014,
97 each lasting three months, starting from January, April, July and October.
98 Each set is composed of 41 seasonal forecast members from which we compute
99 the most likely scenario. This scenario is used as the only forecasted state
100 of the atmosphere. We apply the same methods using the 35 years of ERA-
101 Interim to learn the relation between the surface wind speed and the large-scale
102 circulation of the atmosphere, and apply this relation to the forecasted state
103 of the atmosphere to predict wind speed distribution.

2.2. Methods

At a monthly to seasonal timescale, the surface wind speed is mainly explained by the large scale circulation of the atmosphere. The geopotential height at 500hPa (Z500) and the Mean Sea Level Pressure (MSLP) are variables that well summarize this circulation. In this paper, we only present the results of reconstruction using the Z500 variable as a predictor of the surface wind speed. Indeed, results found when adding MSLP to Z500 predictor were comparable and the improvement was neither systematic nor significant.

In the following paragraphs, we describe in detail the reconstruction methodology which is summarized in Figure 1.

Our attempt aims at reconstructing the distribution of winds on the monthly to seasonal timescales, but not at reconstructing daily timeseries of winds. Indeed, our reconstruction methodology is based on the principal components analysis of the Z500 predictor which informs about the large-scale state of the atmosphere. This knowledge will constrain the likely distribution of surface winds on timescales larger than the lifetime of individual synoptic systems (fronts, storms) and thus will not allow to reconstruct such high frequency timeseries. Following the common practice, we use the Weibull distribution to summarize the surface wind speed distribution [24, 25].

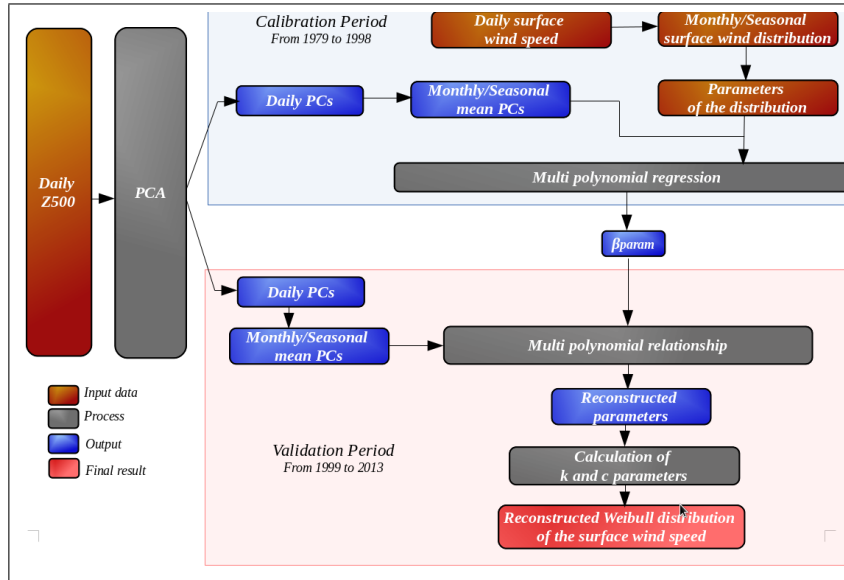


Figure 1: Flow chart describing the reconstruction methodology

124 *Principal component analysis.* To obtain a more compact representation of
 125 the large-scale situation we perform a Principal Component Analysis (PCA)
 126 on Z500. It results in a set of Empirical Orthogonal Functions (EOF), which
 127 represent the typical oscillation patterns spanning the North Atlantic do-
 128 main. Each EOF is associated with one scalar timeseries (the corresponding
 129 PC) which describes how each pattern evolves in time. Figure 2 shows the
 130 five first EOFs and their associated PCs.

131 The first PC corresponds to the seasonal cycle (Fig 2. a,b), explaining
 132 as much as 54.1% of the variance in the dataset: in winter the meridional
 133 pressure gradient strengthens, leading to stronger winds and more intense
 134 synoptic systems. The following four PCs have a clear physical interpretation
 135 [26, 27], they all be related to teleconnection patterns, respectively the North
 136 Atlantic Oscillation (NAO) (Fig 2. c,d), the Eastern Atlantic Pattern (EA)
 137 (Fig 2. e,f), the Scandinavian pattern (SCA) (Fig 2. g,h) and the 2nd
 138 European pattern (EU2) (Fig 2. i,j). These five first PCs explain 76.9% of
 139 the variance in the entire dataset.

140 *Weibull distribution.* To summarize the wind distributions, we choose the
 141 Weibull distribution as the parametric representation for montly and seasonal
 142 distribution of the surface wind speed at a given location. This theoretical
 143 distribution is widely used in the wind energy industry [28, 29, 24]. It pro-
 144 vides a simple way to represent the wind distribution as it is based on only
 145 two parameters: the shape parameter and the scale parameter. We must
 146 highlight the fact that other theorical distributions better capture the shape
 147 of the real wind distribution. In particular, the Rayleigh-Rice distribution
 148 can have two modes, which is not the case for the Weibull [21].

The probability density function (PDF) and the cumulative distribution
 function (CDF) of the Weibull distribution are expressed as follows.

$$f(u; k, c) = \frac{k}{u} \left(\frac{u}{c} \right)^k e^{-(u/c)^k} \quad (1)$$

$$F(u; k, c) = 1 - e^{-(u/c)^k}, \quad (2)$$

149 where u is the wind speed, k and c are respectively the shape and the scale
 150 parameter.

We now define three ways to reconstruct the parameters k and c from the
 data. The WAsP method, referred in the following as WAsP [30], computes
 these parameters from the moments \bar{U} and \bar{U}^3 , as well as the probability of
 exceeding the mean wind speed $1 - P(\bar{U})$ (which must be estimated from the

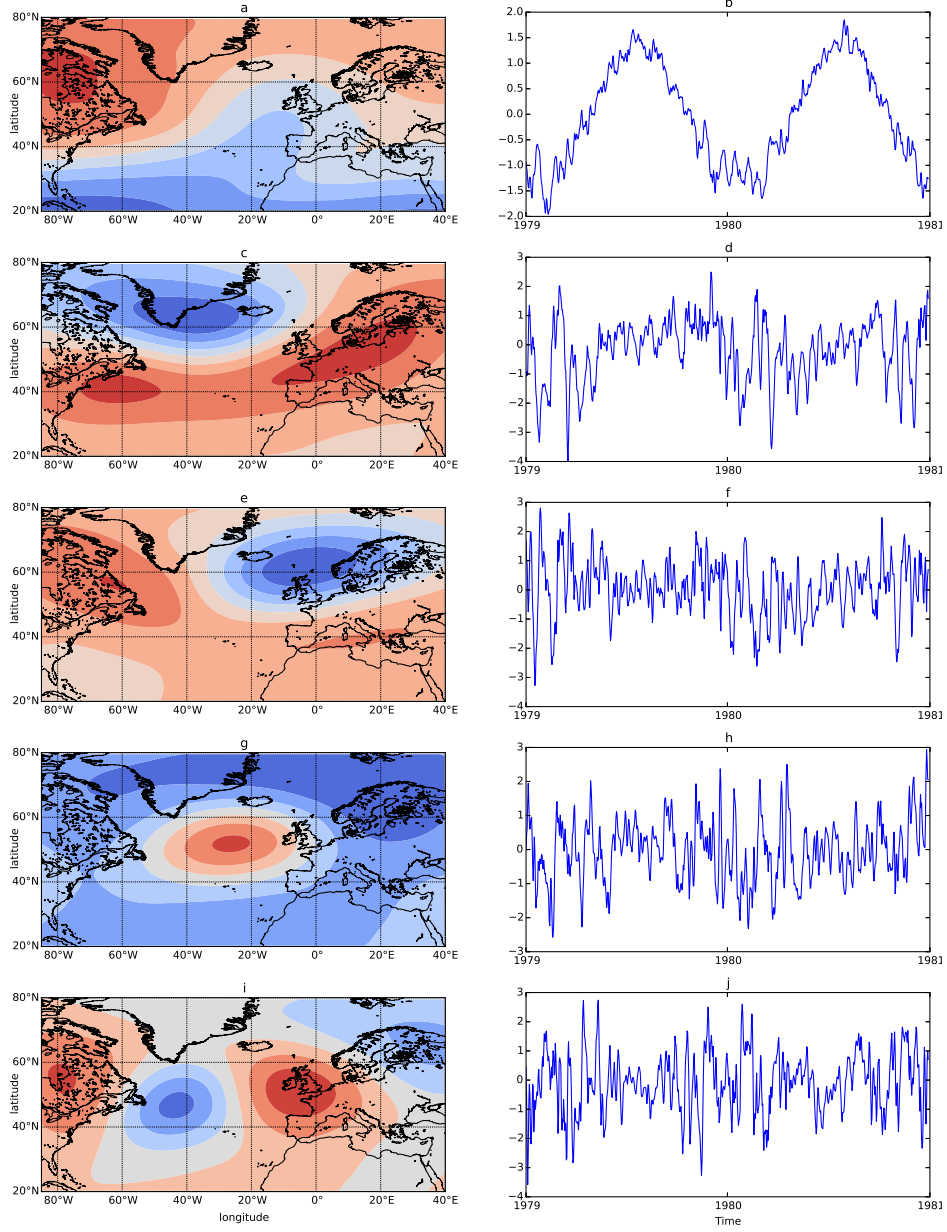


Figure 2: Five firsts EOFs (left side) and five first PCs (right side) of the PCA performed on the 35years and on the entire domain of ERAI Z500 dataset

data). The method focuses on the right-hand tail of the Weibull distribution, which is an important part of the distribution in terms of energy [31]. This

is why the WAsP method is preferred amongst the wind energy industry. In this method, k and c are calculated by solving the following equations.

$$\frac{\overline{U}^3}{\overline{U^3}} \Gamma\left(1 + \frac{3}{k}\right)^{\frac{k}{3}} = -\ln(1 - P(\overline{U})) \quad (3)$$

$$c = \sqrt[3]{\frac{\overline{U^3}}{\Gamma(1 + \frac{3}{k})}} \quad (4)$$

In a second method, referred in the following as KCrec, we take advantage of the fact that the Weibull distribution is given by two parameters, k and c , and straightforwardly reconstruct these: they are fitted by the Maximum Likelihood Estimator (MLE) [32] on the calibration period. The MLE of the Weibull parameters is defined by the following equations.

$$\frac{\sum_{i=1}^n u_i^k \ln(u_i)}{\sum_{i=1}^n u_i^k} - \frac{1}{k} - \frac{1}{n} \sum_{i=1}^n \ln(u_i) = 0, \quad (5)$$

$$c = \frac{\sum_{i=1}^n u_i^k}{n}. \quad (6)$$

A last method was introduced in order to take into account how spread out the wind distribution is. This method, referred in the following as Perc, uses two values, $F(u_1)$ and $F(u_2)$, of the Weibull distribution function, corresponding to wind speeds u_1 and u_2 . The Weibull k and c parameters are then given explicitly by:

$$c = \frac{\ln \ln\left(\frac{1}{1-F(u_2)}\right) \ln(u_1) - \ln \ln\left(\frac{1}{1-F(u_1)}\right) \ln(u_2)}{\ln \ln\left(\frac{1}{1-F(u_2)}\right) - \ln \ln\left(\frac{1}{1-F(u_1)}\right)}, \quad (7)$$

$$k = \frac{c}{u_1} \ln \ln\left(\frac{1}{1-F(u_1)}\right). \quad (8)$$

151 In order to determine the optimal values of u_1 and u_2 , a synthetic test was
 152 performed. First, we generated 30 (one month) or 90 (one season) samples
 153 from the reference Weibull distribution with parameters $k = 2$ and $c = 3.5$.
 154 Next, we determined the two Weibull parameters from the simulated samples
 155 using the Perc method, using different combinations (u_1, u_2) . To find the
 156 best combination, we compared the resulting distributions with the reference
 157 distribution using the Cramer-von Mises (CvM) score (see Appendix). It
 158 was found that the best combination on a monthly scale is the 11th and 83rd

159 percentile. On the seasonal scale, the optimal combination is the 17th and
 160 the 92nd percentile. The combination of the percentiles was not found to be
 161 very sensitive, as there was a small region around the optimum combination
 162 with very similar scores.

163 *Multi-polynomial regression.* We propose to link the large-scale situation
 164 and surface wind speed distribution by a multi-polynomial regression tak-
 165 ing the monthly mean PCs as explanatory variables and the parameters of
 166 the Weibull distribution as dependent variables:

$$\tilde{P} = \beta_0 + \sum_{n=1}^N \beta_{n,n} C_n(t)^2 + \sum_{n=1}^{N-1} \sum_{m=n+1}^N \beta_{n,m} C_n(t) C_m(t). \quad (9)$$

167 Here, \tilde{P} is the dependent variable (Weibull parameter k or c for a given loca-
 168 tion), C_n are the principal components and $\beta_{n,n}$ and $\beta_{n,m}$ are the regression
 169 weights found by least squares. The number N of principal components is
 170 determined by cross validation as explained below. We perform the regres-
 171 sion on a calibration period of 20 years between 1979 and 1998. This results
 172 in weights quantifying the relationship between the large-scale circulation
 173 and the Weibull parameters for each individual location. These weights can
 174 be combined with the known PC values on the reconstruction period of 15
 175 years between 1999 and 2013 to reconstruct the monthly/seasonal Weibull
 176 distribution.

177 *Optimizing the number of principal components through cross validation.* The
 178 first five PCs of the Z500 can be easily interpreted as predictors of the wind.
 179 Still, to a certain extent, the following PCs can also explain the variability of
 180 the wind at the monthly/seasonal scale. To check whether taking five PCs is
 181 really optimal, we performed a cross-validation procedure. For this purpose,
 182 we calculated the temporally and spatially averaged CvM score (see Ap-
 183 pendix) of 7 reconstructions of 5 years each, taking the remaining 30 years of
 184 the data set as calibration period. Figure 3 plots the CvM scores as function
 185 of the number of PCs used. The minimum mean CvM is clearly apparent
 186 for all three methods for both monthly (Fig 3. a, b, c) and seasonal (Fig 3.
 187 d, e, f,) reconstruction. This minimum is around five PCs which confirms
 188 the fact that the large-scale circulation variability is accurately linked to the
 189 wind speed variability at the monthly and seasonal timescale.

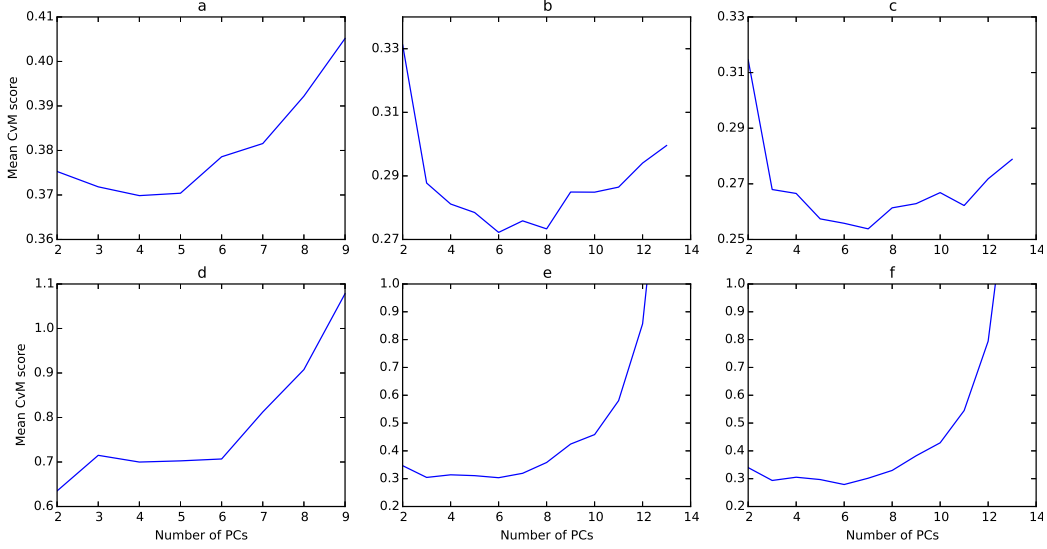


Figure 3: Mean CvM score obtained by cross validation in function of the number of PCs used to reconstruct the distribution of the surface wind speed. From left to right: Wasp (a,d), Perc (b,e), and KCreC (c,f) methods; top: CvM score for monthly wind distribution reconstruction (a,b,c); bottom: CvM score for seasonal wind distribution reconstruction (d,e,f)

3. Evaluating the reconstruction methods

As mentionned in the Introduction, we use the wind speed from the ERAI-reanalysis as the reference wind speed. To assess the reconstruction quality, the CvM score (see Appendix) is calculated between the reconstructed CDF and the real wind CDF. The CvM scores of the reconstructed wind speed distributions are then compared to the CvM scores computed between the real wind distributions and the climatological distributions. In simple terms, the climatological distribution is the distribution of all values of wind for each month or season in one specific location, based on all reanalysis data from this location and the specific month or season. The climatological distributions are usually used by the industry to have a first assessment of the wind energy production at a seasonal time scale. An example of real, climatological and reconstructed wind speed CDFs is shown in Figure 4.

3.1. Performance of methods for wind speed distribution reconstruction

The CvM score allows to test the null hypothesis (H_0) that the two samples come from the same distribution. Assuming that the reconstructed

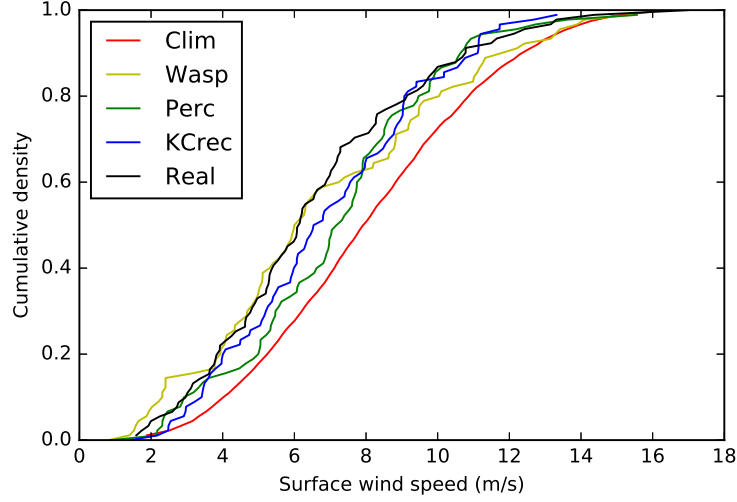


Figure 4: Real, climatological, and reconstructed seasonal CDFs for winter 2012 at 48.5°N 3.0°W

206 distributions and the real distributions are based on samples large enough to
 207 say that the corresponding CvM scores follow the limiting distribution, we
 208 can define the p-value corresponding to 95% confidence (see Appendix). If
 209 the calculated CvM score is below this value, we can say at 95% confidence
 210 that the two compared samples come from the same distribution. We com-
 211 pare results of the tests for the climatology and the reconstruction methods.
 212 We can define five different cases:

- 213 • Case A: H_0 is not rejected for the method and rejected for the clima-
 214 tology
- 215 • Case B: H_0 is not rejected for both and the CvM of the method is
 216 smaller than the CvM of the climatology
- 217 • Case C: H_0 is not rejected for both and the CvM of the method is
 218 larger than the CvM of the climatology
- 219 • Case D: H_0 is rejected for the method and not rejected for the clima-
 220 tology
- 221 • Case E: H_0 is rejected for both the method and the climatology

Results over the whole domain in all different cases are given in table 1 and 2 for monthly and seasonal reconstruction respectively. We compare the reconstruction methods to not only the classical climatology (a), but also to the parametric climatology (b).

Indeed, the hypothesis of the Weibull distribution introduces a bias in the distribution reconstruction which is not present in the classical climatology. In order to have a fair comparison, we also fit by MLE a Weibull distribution on the historical data referred as the parametric climatology.

Methods	Wasp		Perc		KCrec		Clim	Parametric Clim
CvM < p	69.1		82.1		85.2		89.3	81.7
Comparison with	a	b	a	b	a	b	-	-
Case A	5.8	11.5	6.5	11.9	6.9	13.0	-	-
Case B	17.8	24.1	25.0	34.0	27.3	37.1	-	-
Case C	45.5	33.5	50.5	36.2	51.1	35.1	-	-
Case D	26.0	24.1	13.8	11.6	11.2	9.5	-	-
Case E	4.8	6.8	4.1	6.3	3.7	5.3	-	-

Table 1: Percentage of time the result of the CvM test gives Cases A,B,C,D, or E on the whole domain, for the entire validation period, for monthly reconstructed distribution compared to the classical climatology (a) and to the parametric climatology (b). The p-value, p, is 0.46136 for 95% confidence level (see appendix)

Methods	Wasp		Perc		KCrec		Clim	Parametric Clim
CvM < p	44.3		73.3		79.8		88.6	77.3
Comparison with	a	b	a	b	a	b	-	-
Case A	3.8	8.9	5.5	11.0	6.1	13.9	-	-
Case B	10.5	13.6	22.8	31.2	23.8	34.2	-	-
Case C	30.0	21.9	45.0	31.1	49.9	31.7	-	-
Case D	48.1	41.9	20.8	15.0	14.9	11.4	-	-
Case E	7.5	13.8	5.9	11.7	5.3	8.8	-	-

Table 2: Same as table 1 but for seasonal distribution

The first lines of tables 1 and 2 show the fraction of time each method gives a distribution not discernable from the real distribution at 95% confidence level. It shows that all methods, appart from Wasp, have a good ability to reconstruct the real wind distribution. We can also see that fitting a Weibull distribution on the climatology reduces by about 10% this percentage. Cases A and B summarize the number of time each method is

236 doing better than the climatology (non-parametric (a) or parametric (b)).
 237 On the contrary, Cases C and D summarize the number of time the cli-
 238 matology is doing better than the method. On average, on the all domain
 239 and for monthly and seasonal timescales, the non-parametric climatology (a)
 240 do better than every methods more than 60% of the time (78.1% against
 241 Wasp at the seasonal scale, to 62.3% against KCrec at the monthly scale).
 242 Nevertheless, when comparing to the parametric climatology, for monthly
 243 and seasonal reconstruction, the KCrec method performs 49.1% of the time
 244 better at monthly scale, and 48.1% at the seasonal scale. This shows again
 245 the error brought by the Weibull distribution reconstruction. In all cases,
 246 methods perform better at the monthly scale than at the seasonal scale. It
 247 is interesting to notice that the cases for which the percentage is increased
 248 at the seasonal scale are cases D and E, corresponding to times when recon-
 249 structed distribution cannot be believed to come from the same distribution
 250 as the real sample, at 95% confidence level. (Tables 1 and 2).

251 Figure 5 and 6 show on average on the validation the number of time
 252 each method behaves better than the classical climatology (Cases A and B).
 253 It can be seen that the Perc and KCrec methods do better than the Wasp
 254 method. Indeed, at monthly timescale, the Perc and KCrec methods can
 255 do better than the climatology in average more than 30% of times, while
 256 the Wasp method does better than the climatology about 25% of times on
 257 average displaying a clear difference between north and south (Figure 5 and
 258 Table 1). On a seasonal scale, the Wasp method performs clearly worse
 259 than at a monthly scale. The Perc and KCrec methods at a seasonal scale
 260 display an interesting spatial variability. Indeed, they do more than 40%
 261 of times better than the climatology in the north of France, whereas in the
 262 south, this percentage is about 20% to 25% (Figure 6). When comparing
 263 to the parametric climatology, all methods display the same pattern, but all
 264 percentages are increased more than 10% (Not shown).

265 We can argue that the climatology does not reproduce well the extremes
 266 of the wind distribution that is to say the strongest winds because it acts as
 267 a filter of high frequency wind variations. In the northern part of France,
 268 the storm track in winter and autumn brings stronger winds than in spring
 269 and summer. We can assume that the reconstruction methods based on the
 270 PCs of Z500 may better reproduce those strong winds than the climatology,
 271 because the storm track position and strength is mainly driven by the NAO
 272 and SCA oscillation patterns. Figure 7 shows the ratio of the number of times
 273 each method is doing better than the climatology for seasonal distributions,

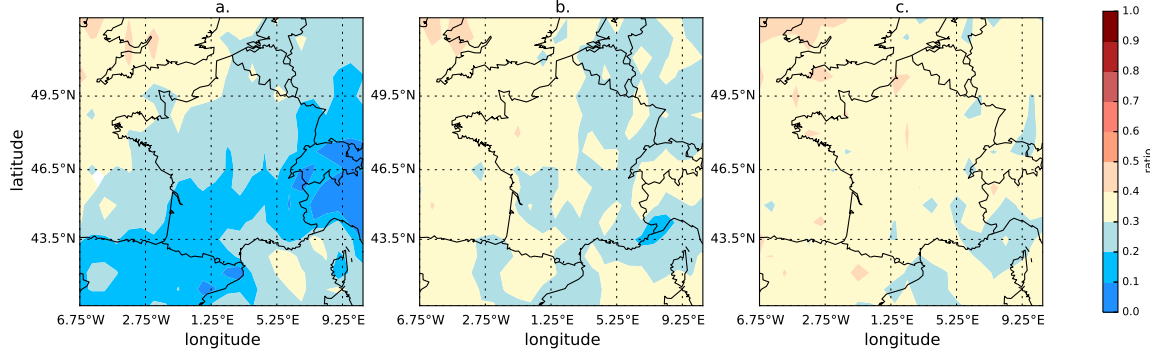


Figure 5: Fraction of times each method does better than the climatology (cases A and B) for monthly distribution reconstruction. From left to right: Wasp (a), Perc (b), KCrec (c)

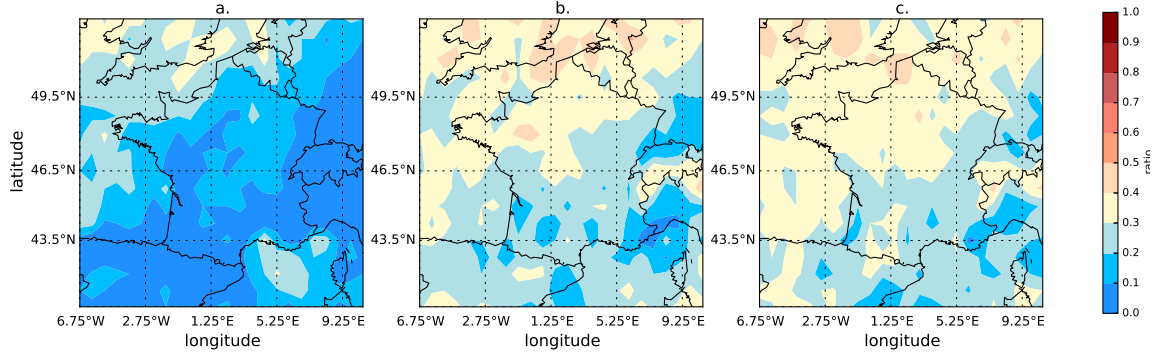


Figure 6: Same as Figure 5 but for seasonal distribution reconstruction.

by taking each season separately. We can clearly see on this figure that the performance regarding the climatology of the Perc and KCrec methods, and to a certain extent the Wasp method, depends on the season and on the region. Indeed, both the Perc and the KCrec methods display a high percentage of times (up to 70% at some points) when they do better than the climatology in the north of France for the winter and autumn seasons.

3.2. Performance of the methods for estimating the capacity factor

For wind energy purposes, it is not exactly the full wind distribution that needs to be estimated. For a given turbine, once the wind is between the nominal wind speed and below the cut-out speed, the precise value does not matter. In the present section we take this into account and reevaluate each

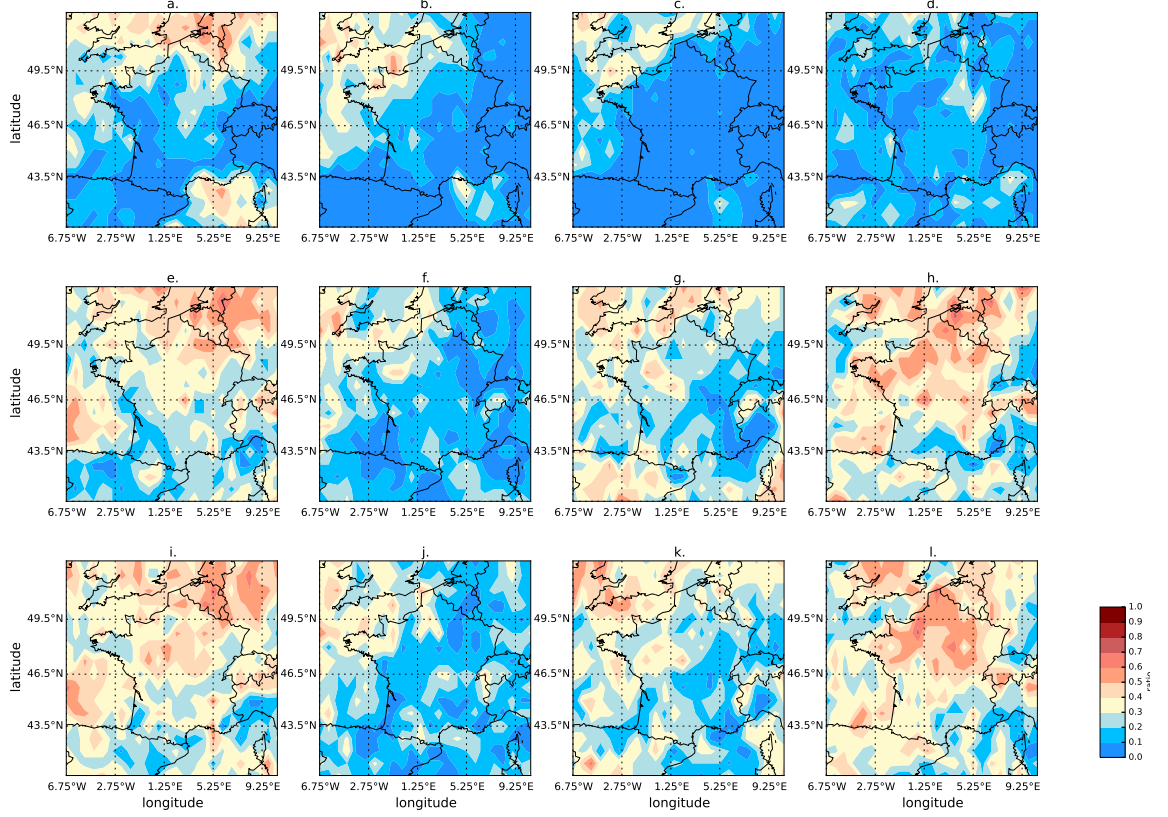


Figure 7: Fraction of time each method do better than the climatology (cases A and B) for seasonal distribution reconstruction based on Z500 for each season. From left to right: Winter, Spring, Summer and Autumn; From top to bottom: Wasp, Perc, and KCreC methods

method. A preliminary step consists in designing a procedure which mimicks the weighting of wind values by a power curve, in a manner which accounts for the considerable geographical variations of the wind (a single, generic power curve would not make sense).

Each wind turbine is characterized by its power curve which gives the output power as function of the wind speed. The energy produced during a given period can be expressed as :

$$E = T \int_0^{\infty} P_{out}(u) dU, \quad (10)$$

where T is the period considered (month or season) and $P_{out}(u)$ is the output power given the wind speed u . The capacity factor is defined as the ratio

294 between the actual energy produced during a given period and the energy
 295 that would have been produced if the wind turbine had run at its maximum
 296 power during the entire period :

$$CF = \frac{E}{P_n T}, \quad (11)$$

297 where P_n is the nominal power of the wind turbine.

298 In order to take into account the fact that the data used are at 10-meter
 299 height and the mean wind speed is highly varying among different locations,
 300 we use a location-adapted power curve, proposed by Jourdier [20]. In this
 301 curve, the wind speed is divided by a location-dependent parameter a , cho-
 302 sen so that the modified power curve has a capacity factor of 23% on the
 303 calibration period. This corresponds to the average capacity factor in France
 304 in 2014 [33]. This procedure is illustrated in Figure 8.

305 To assess the accuracy of the reconstructed capacity factor, the relative
 306 error between the reconstructed capacity factor and the capacity factor from
 307 the reanalysis is computed :

$$\Delta CF = \frac{CF - CF_{real}}{CF_{real}} \quad (12)$$

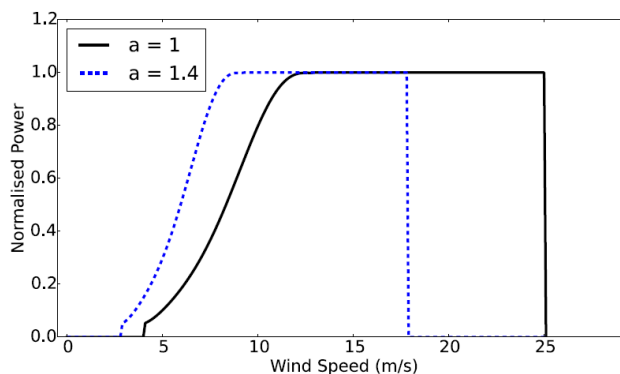


Figure 8: Example of the location-adapted power curve. In solid black: the real power curve for wind speed at 80m height; in dashed blue: the adapted power curve. It has the same shape, but the wind speed is divided by a number a to achieve a capacity factor of 23%.

308 Figures 9 and 10 show the relative error on the calculated capacity factor
 309 for monthly and seasonal reconstructions respectively. At both timescales,

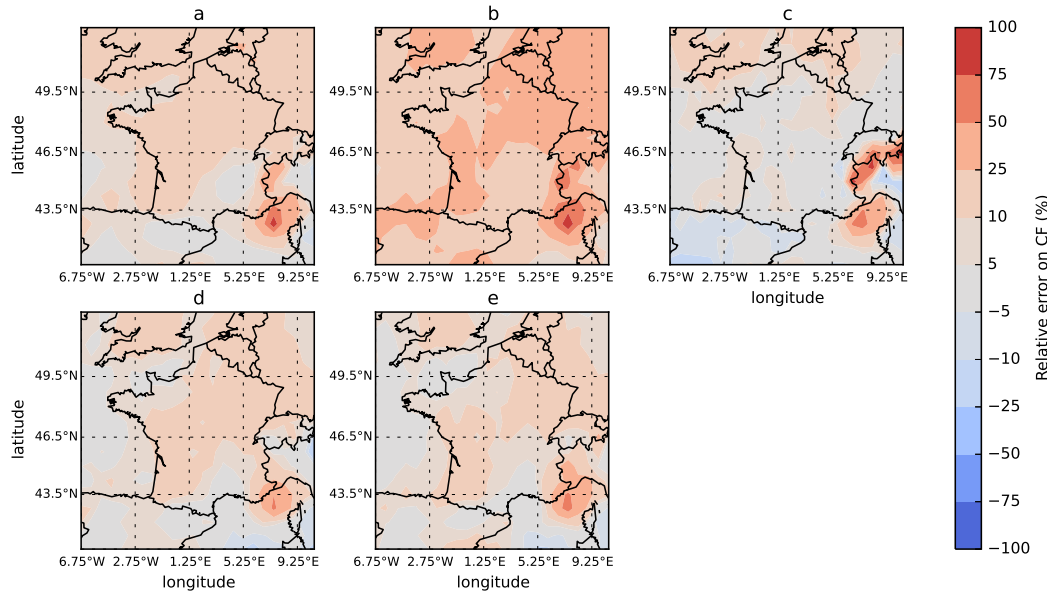


Figure 9: Relative error on the capacity factor (%) for monthly distributions given by: non parametric climatology (a), parametric climatology (b), Wasp (c), Perc (d), and KCreC (e)

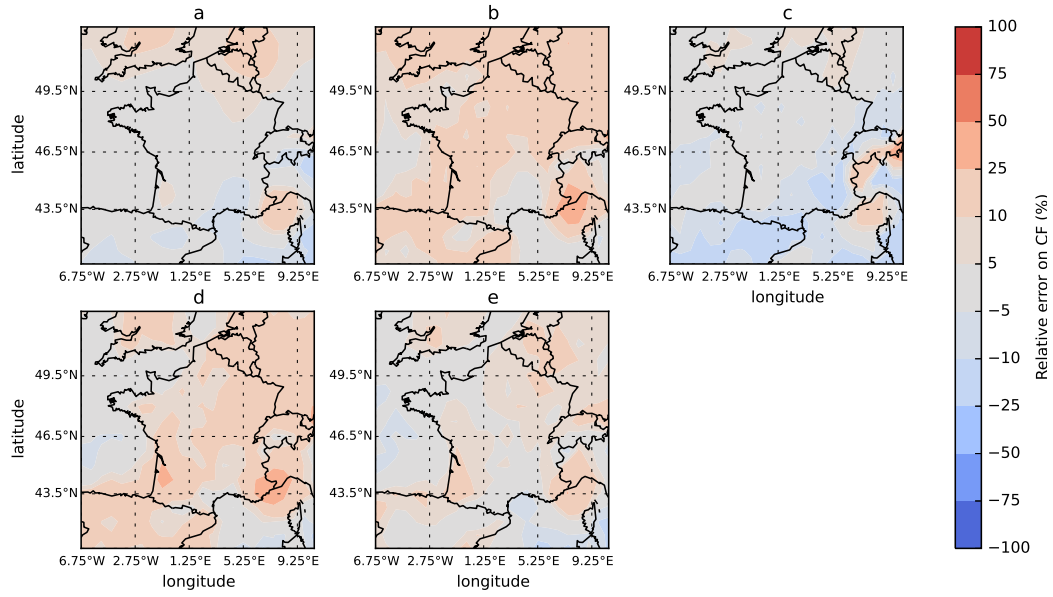


Figure 10: Same as Figure 9 but for seasonal distributions.

310 the Perc method overestimates it mostly onshore by about 25% on average.
 311 The KCrec method behaves like the Perc methods at the monthly scale, but
 312 is performing better at the seasonal scale with an overestimation of about
 313 10% onshore. As expected, the Wasp method shows good performance in es-
 314 timating the capacity factor as its reconstruction focuses on the right tail of
 315 the Weibull distribution. Nevertheless, it overestimates the energy produc-
 316 tion in the northern part of France at a monthly scale and underestimates
 317 it in the southern part of France at a seasonal scale. The non-parametric
 318 climatology behaves very well at the seasonal scale even though it displays a
 319 slight overestimation in the north of France. At the monthly scale, on aver-
 320 age, on the entire domain it overestimates the capacity factor by about 25%.
 321 By contrast, the parametric climatology behaves very badly at the monthly
 322 scale, overestimating the energy production by 50% in average. At a seasonal
 323 scale, this overestimation decreases but is still high, highlighting again the
 324 error induced by the Weibull distribution hypothesis.

325 In any case, there is a tendency of all methods to overestimate the capacity
 326 factor, mostly onshore. The climatology acts as a filter of high frequency
 327 variation of the wind, meaning that it does not describes well the tails of
 328 the distribution. As the power curve is designed so that the wind turbine
 329 works at its nominal power near the mean wind speed, this results in an
 330 overestimation of the capacity factor.

331 On the other hand, Drobinski et al. [21] showed that a Weibull distri-
 332 bution fitted by MLE describes well the center of the distribution (near the
 333 mean wind speed), but tends to underestimate the tails of the distribution.
 334 This leads to the same consequence. That explains why the parametric clima-
 335 tology acts worse than the non-parametric climatology, but also why KCrec
 336 overestimates the capacity factor. This has no such effect offshore because
 337 the wind above sea is steadier so that the distribution is more peaked around
 338 the mean. Regarding the Perc method, the Weibull reconstruction is based
 339 on two percentiles defined to minimize the CvM score. It may results in
 340 the same effect of underestimation of the tails of the distribution. Future
 341 work could focus on a sensitivity analysis to the percentiles definition by
 342 minimizing the error on capacity factor.

343 At the seasonal scale, the real distribution is based on a larger sample
 344 which implies that the center of the distribution has a much larger weight
 345 than the tails at this scale than at the monthly scale. The effect of underes-
 346 timating the tails is thus less visible.

4. Towards monthly and seasonal forecast of the wind speed distribution

The analysis described above has shown that the large-scale state of the atmosphere contains information on the likely distribution of surface winds, and our proposed methods allow to recover at least part of this information. A long-term perspective will be to use this to build forecasts of surface wind distributions. Below we present a preliminary attempt based on existing seasonal forecasts, to assess the potential of this method for monthly or seasonal forecasts.

A first step is to assess the skill in seasonal forecasts for predicting the large-scale state of the atmosphere in our region of interest. The root mean square error (RMSE) between the daily PCs of Era-Interim and those of the seasonal forecast is shown in Figure 11. This figure gives an idea of the lead-time of such a forecast. It shows that the error increases rapidly until it levels off after 20 days indicating that there is no more valuable information on the large-scale circulation in the data. As a consequence, it will not be possible to have an accurate wind distribution forecast at more than the monthly horizon.

One technical difficulty arises: the monthly distribution of wind coming from the ECMWF analysis stands for the real distribution. As the analysis does not come from the same model as the ERA-Interim data, a bias exists between the distributions coming from the analysis and the distributions based on ERA-Interim data. We thus apply a classical quantile/quantile correction between the 4 years based distributions of the analysis and of ERA-Interim between 2012 and 2015 at each point of the gridded domain. We apply this correction to the monthly wind distribution of the analysis. Because of the small amount of forecasts and of the uncertainties due to the bias, we will not be able to have the same deep analysis as in the reconstruction part of the paper. The corrected monthly distribution of the wind speed coming from the analysis is compared to the climatology of ERA-Interim and to the forecast distributions using the CvM score.

The percentage of time each method does better than the climatology, averaged over the entire domain, for the 1st month of the 12 forecasts, is summarized in table 3. The results for the Perc and KCrec methods are comparable to the reconstruction results. On the contrary, the Wasp method shows a very high score when evaluating the entire distribution and a lesser score when evaluating the energy production, which is not consistent with

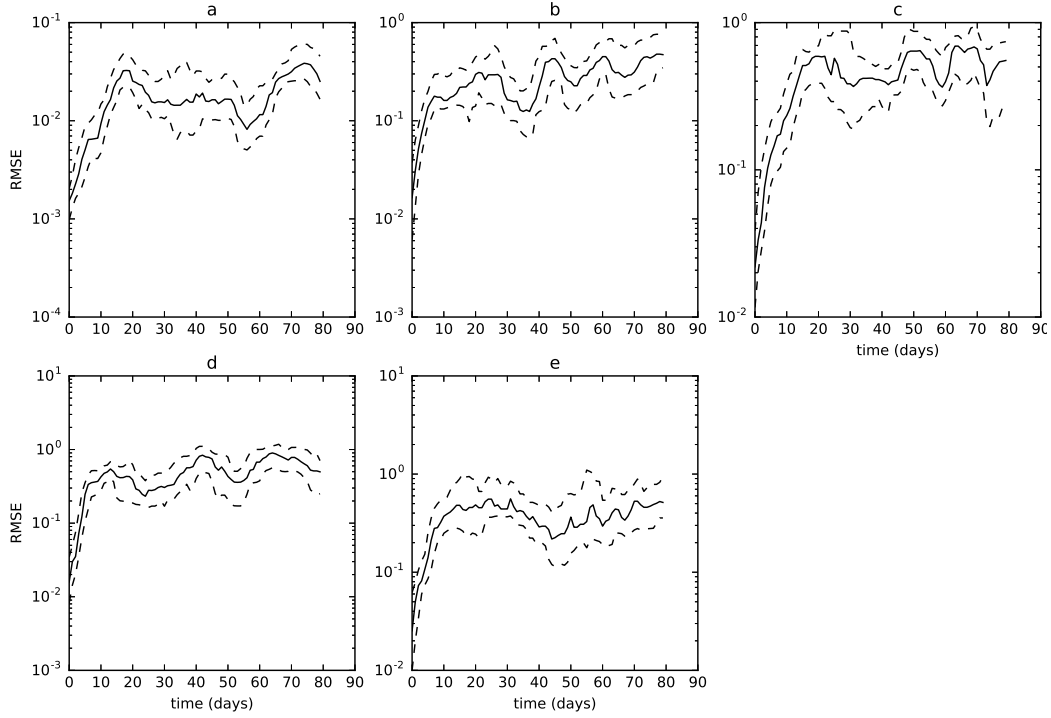


Figure 11: RMSE calculated between the PCs of Era-Interim and the PCs of the seasonal forecast. The solid line represents the median of the error, dashed lines represent the 60th percentile (top) and the 40th percentile (bottom). a. Seasonal, b. NAO, c. EA, d. SCA, e. EU2

Forecast method	Wasp	Perc	KCrec
total 1st month	46.4 (31.2)	20.2 (25.5)	28.8 (27.5)
2012	41.0 (35.0)	15.1 (20.5)	22.9 (23.6)
2013	44.1 (25.7)	22.9 (25.4)	32.4 (26.5)
2014	54.0 (33.0)	22.5 (30.4)	31.3 (32.3)

Table 3: Percentage of the number of times each method does better than the climatology on the whole domain for the 3 years of forecasts. First values correspond to the evaluation of the entire distribution; values in parenthesis corresponds to the evaluation of the distribution between the cut in and the cut out.

384 the reconstruction results. When calculating the error on the capacity factor,
385 the forecast methods always highly overestimate the wind energy production
386 onshore (more than 100% at some points), and slightly underestimate it off-
387 shore (more than 10%). The non-parametric climatology overestimates the

388 capacity factor by more than 10% onshore and underestimates it offshore,
389 whereas the parametric climatology highly overestimates the energy produc-
390 tion on the whole domain as it was the case in the evaluation part.

391 Regarding the large uncertainty due to the limited number of forecasts,
392 the robustness can be inferred from the consistency of the forecasts results
393 with those obtained in the previous section.

394 Still, work must be continued to evaluate the forecasts performance of
395 such methods, by using larger sets of numerical seasonal weather forecast,
396 but also by testing methods based on non-parametric distribution estimation.

397 5. Conclusion

398 In this paper, a new approach for modelling the wind speed at the seasonal
399 scale has been proposed. We suggest to model not only the mean wind speed
400 but the entire monthly/seasonal distribution of the wind. Linking the wind to
401 its synoptic predictors we have shown that there is valuable information in the
402 large-scale circulation variability that can explain the wind speed distribution
403 at such long timescales. The proposed methods show good performances in
404 reconstructing the monthly and seasonal wind speed distributions even if
405 the climatology is still a good predictor. Moreover, reconstruction methods
406 performances display an interesting spatial and seasonal variability. Indeed,
407 in the north of France in winter and fall, the proposed methods showed
408 better ability to model strong winds than the climatology. Nevertheless,
409 the attempt of forecasting also highlights the fact that seasonal forecasts of
410 ECMWF are not yet mature enough to give valuable information on the
411 large-scale circulation variability at the horizons exceeding a month.

412 Acknowledgments

413 This research was supported by the ANR project FOREWER (ANR-14-
414 CE05- 0028). This work also contributes to the HyMeX program (HYdro-
415 logical cycle in The Mediterranean EXperiment [34]) through the working
416 group Renewable Energy.

417 Appendix: Cramer-Von Mises score

418 To assess the reconstruction quality, we use the Cramer-Von-Mises score
419 defined in Anderson et al. [35]:

$$CvM = \frac{MN}{M+N} \int_{-\infty}^{\infty} [F_N(x) - F_M(x)]^2 dH_{M+N}(x) \quad (.1)$$

420 Here, M and N are the sample sizes in each of the distributions, $F_N(x)$
421 and $F_M(x)$ are the CDFs of the two samples and $H_{M+N}(x)$ is the combined
422 distribution of the two samples together. The smaller the CvM score, the
423 better the goodness of fit between the two tested distributions. Anderson et
424 al. [35] showed that Equation (.1) is equivalent to

$$CvM = \frac{U}{NM(M+N)} - \frac{4NM-1}{6(N+M)}, \quad (.2)$$

425 where $U = N \sum_{i=1}^N (r_i - i)^2 + M \sum_{j=1}^M (r_j - j)^2$, r_i are the ranks of the elements
426 of the sample of size N in the combined sample and r_j are the ranks of the
427 sample of size M in the combined sample.

428 The CvM score allows to test the null hypothesis H_0 : "the two samples
429 come from the same distribution". When $M \rightarrow \infty$ and $N \rightarrow \infty$, under the null
430 hypothesis, the CvM score follows the limiting distribution with mean $\frac{1}{6}$ and
431 variance $\frac{1}{45}$. In this configuration, the p-value giving 95% confidence that
432 the null hypothesis is true is $p = 0.46136$, [35].

433 References

- 434 [1] EWEA, Wind in power: 2014 European statistics, European Wind En-
435 ergy Association.
- 436 [2] S. S. Soman, H. Zareipour, O. Malik, P. Mandal, A review of wind power
437 and wind speed forecasting methods with different time horizons, North
438 American Power Symposium (NAPS) (2010) 1–8.
- 439 [3] W. Chang, A literature review of wind forecasting methods, Journal of
440 Power and Energy Engineering 2 (2014) 161–168.
- 441 [4] A. Sfetsos, A novel approach for the forecasting of mean hourly wind
442 speed time series, Renewable Energy, 27 (2002) 163–174.

- 443 [5] P. Gomes, R. Castro, Wind speed and wind speed forecasting using sta-
444 tistical models: Autoregressive Moving Average (ARMA) and Artificial
445 Neural Networks (ANN), International Journal of Sustainable Energy
446 Development (IJSED) 1.
- 447 [6] A. Carpinone, M. Giorgio, R. Langella, A. Testa, Markov chain model-
448 ing for very-short-term wind power forecasting, Electric Power Systems
449 Research 122 (2015) 152 – 158.
- 450 [7] J. Taylor, P. McScharry, R. Buizza, Wind power density forecasting
451 using ensemble prediction and time series model, IEEE Transactions on
452 Energy Conversion 34.
- 453 [8] M. Wytock, J. Z. Kolter, Large-scale probabilistic forecasting in energy
454 systems using sparse gaussian conditional random fields, Proceedings of
455 the IEEE Conference on Decision and Control (2013) 1019–1024.
- 456 [9] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, P. S. Dokopoulos,
457 Long-term wind speed and power forecasting using local recurrent neural
458 network models, IEEE Transaction on Energy Conversion 21 (2006) 273–
459 284.
- 460 [10] J. Najac, J. Boe, L. Terray, A multi model ensemble approach for as-
461 sessment of climate change impact on surface winds in France, Climate
462 Dynamics 32 (2009) 615–634.
- 463 [11] D. J. Sailor, M. Smith, M. Hart, Climate change implications for wind
464 power resources in the northwest united states, Renewable Energy 33
465 (2008) 2393–2406.
- 466 [12] S. Pryor, R. Barthelmie, Climate change impacts on wind energy: A
467 review, Renewable and Sustainable Energy Reviews 14 (2010) 430–437.
- 468 [13] M. Bilgili, B. Sahin, A. Yasar, Application of artificial neural networks
469 for the wind speed prediction of target station using reference stations
470 data, Renewable Energy 32 (2007) 2350–2360.
- 471 [14] H. B. Azad, S. Mekhilef, V. G. Ganapathy, Long-term wind speed fore-
472 casting and general pattern recognition using neural networks, IEEE
473 Transaction on Sustainable Energy 5 (2014) 546–553.

- 474 [15] J. Wang, S. Qin, Q. Zhou, H. Jiang, Medium-term wind speeds forecast-
475 ing utilizing hybrid models for three different sites in xinjiang, china,
476 Renewable Energy 76 (2015) 91–101.
- 477 [16] Z. Guo, W. Zhao, H. Lu, J. Wang, Multi step forecasting for wind speed
478 using a modified EMD based artificial neural network model, Renewable
479 Energy 37 (2012) 241–249.
- 480 [17] A. More, M. Deo, Forecasting wind with neural networks, Marine Struc-
481 tures 16 (2003) 35–49.
- 482 [18] J. Owen, T. Palmer, The impact of El Nino on an ensemble of extended
483 range forecasts, American Meteorological Society 115 (1987) 2103–2117.
- 484 [19] C. Cassou, Intraseasonal interaction between Madden-Julian oscillation
485 and the North Atlantic Oscillation, Nature 455 (2008) 523–597.
- 486 [20] B. Jourdier, Wind resource in metropolitan france: assessment methods,
487 variability and trends, Ph.D. thesis, Ecole Polytechnique (2015).
- 488 [21] P. Drobinski, C. COulais, B. Jourdier, Surface wind-speed statistic mod-
489 elling: Alternatives to the Weibull distribution and performance evalu-
490 ation, Boundary-Layer Meteorol 157 (2015) 97123.
- 491 [22] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli,
492 S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer,
493 P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann,
494 C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B.
495 Healy, H. Hersbach, E. V. Holm, L. Isaksen, P. Kallberg, M. Kohler,
496 M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K.
497 Park, C. Peubey, P. de Rosnay, C. Tavorato, J.-N. Thepaut, F. Vitart,
498 The era-interim reanalysis: conguration and performance of the data
499 assimilation system, Q. J. R. Meteorol. Soc. 137 (2011) 553597.
- 500 [23] F. Molteni, T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Fer-
501 ranti, L. Magnusson, K. Mogensen, T. Palmer, F. Vitart, The new ecmwf
502 seasonal forecast system (system 4), ECMWF Technical Memorandum
503 656.
- 504 [24] T. Burton, N. Jenkins, D. Sharpe, E. Bossanyi, Wind energy handbook,
505 Wiley.

- 506 [25] J. Manwell, J. McGowan, , A. Rogers, Wind energy explained. theory,
507 design and application, Wiley.
- 508 [26] M. Vrac, P. V. Ayar, P. Yiou, Trends and variability of seasonal weather
509 regimes, International Journal of Climatology.
- 510 [27] C. Cassou, L. Terray, J. W. Hurrell, C. Deser, North atlantic winter
511 climate regimes: Spatial asymmetry, stationarity with time, and oceanic
512 forcing, American Meteorological Society.
- 513 [28] I. Lun, J. Lam, A study of Weibull parameters using long-term wind
514 observation, Renewable Energy 20 (2000) 145–153.
- 515 [29] C. Justus, W. Hargreaves, A. Yalcin, Nationwide assessment of potential
516 output from wind-powered generators, Journal of Applied Meteorology
517 15 (1976) 673–678.
- 518 [30] I. T. N.G. Mortensen, L. Landberg, E. Petersen, Wind atlas analysis
519 and application program (wasp), vol.1: Getting started. vol.2: Users
520 guide., Ris National Laboratory.
- 521 [31] S. Pryor, M. Nielsen, R. Barthelme, J. Mann, Can satellite sampling
522 of offshore wind speeds realistically represent wind speed distributions?
523 Part II: Quantifying uncertainties associated with distribution fitting
524 methods, American Meteorological Society.
- 525 [32] A. Cohen, Maximum likelihood estimation in the Weibull distribution
526 based on complete and censored samples, Technometrics 7 (1965) 579–
527 588.
- 528 [33] RTE, Syndicat des Energies Renouvelables, ERDF and ADEef,
529 Panorama de l’électricité renouvelable 2014.
- 530 [34] P. Drobinski, V. Ducrocq, P. Alpert, E. Anagnostou, K. Branger,
531 M. Borge, I. Braud, A. Chanzy, S. Davolio, G. Delrieu, C. Estournel,
532 N. F. Boubrahmi, J. Font, V. Grubisic, S. Gualdi, V. Homar, B. Ivancan-
533 Picek, C. Kottmeier, V. Kotroni, K. Lagouvardos, P. Lionello, M. Llasat,
534 W. Ludwig, C. Lutoff, A. Mariotti, E. Richard, R. Romero, R. Rotunno,
535 O. Roussot, I. Ruin, S. Somot, I. Taupier-Letage, J. Tintore, R. Uijlen-
536 hoet, H. Wernli, A 10-year multidisciplinary program on the Mediter-
537 ranean water cycle, Meteorol. Soc. 95 (2014) 1063–1082.

- ⁵³⁸ [35] T. Anderson, On the distribution of two sample Cramer Von Mises
⁵³⁹ criterion, The Annals of Mathematical Statistics (1962) 1148–1159.